**STARBUCKS REPORT**

**Exploratory Data Analysis Report**

**1. Introduction**

This report provides an exploratory data analysis (EDA) of the Starbucks dataset. The objective is to uncover insights, understand patterns, and prepare the data for further analysis or modeling. The dataset includes customer reviews, ratings, and associated metadata.

**2. Objectives**

* Understand the structure and content of the dataset.
* Identify and address data quality issues.
* Analyze variables to derive insights using statistical and visualization techniques.

**3. Research Methodology**

* **Data Cleaning**: Missing values, duplicates, outliers, and incorrect data types were handled.
* **Statistical Analysis**: Summary statistics and correlation analysis were performed.
* **Visualization**: Data patterns were explored using univariate, bivariate, and multivariate visualizations.

**4. Data and Variables**

* **name**: Name of the reviewer.
* **location**: Location of the Starbucks branch.
* **Date**: Date of the review.
* **Rating**: Customer rating (numerical).
* **Review**: Textual feedback from the customer.
* **Image\_Links**: Links to images included with the reviews.

**5. Statistical Tools and Techniques**

* **Descriptive Analysis**: Used summary statistics (mean, median, mode, standard deviation) for univariate analysis.
* **Correlation Matrix**: Analyzed relationships among numerical variables.
* **Visualization Techniques**:
  + Histograms and bar charts for distribution analysis.
  + Box plots for location-based rating comparisons.
  + Scatter plot matrix for multivariate analysis.

**6. Findings**

* **Data Quality**:
  + Missing values and duplicates were identified and removed.
  + The date column was successfully converted into a proper date format.
  + Outliers in ratings were treated using the interquartile range (IQR) method.
* **Univariate Analysis**:
  + Ratings are mostly concentrated around high values, as shown in the histogram.
  + A bar chart highlights the frequency of individual ratings.
* **Bivariate Analysis**:
  + Box plots revealed variation in ratings across different locations.
  + Ratings were found to be generally consistent across locations.
* **Multivariate Analysis**:
  + Scatter plot matrix revealed no significant anomalies among numerical variables.
  + Correlation analysis showed strong relationships between specific numerical features.

**7. Conclusion**

This EDA has highlighted key patterns in the Starbucks dataset. The data is clean and ready for further analysis or modeling. Insights gained include a generally high customer satisfaction (ratings) and notable consistency across different locations.

**8. Recommendations**

* Perform sentiment analysis on customer reviews for deeper insights.
* Investigate specific branches with lower ratings for targeted improvements.
* Use the cleaned dataset for predictive modeling to forecast customer ratings.